ELSEVIER

PII: S0968-5677(98)00126-6

Supramolecular Science **5** (1998) 787–790

0968-5677/98/$19.00

Conformation-driven computing: a comparison of designs based on DNA, RNA, and protein

Michael Conrad and Klaus-Peter Zauner
Wayne State University, Department of Computer Science, Detroit, Michigan 48202, USA

Molecular pattern recognition based on conformational interactions is the major basis of control and information processing in biological cells. Designs for pattern recognition devices that use the conformational characteristics of different major classes of biomolecules (DNA, RNA, and protein) are considered.

(Keywords: molecular computing; biological information processing; cellular control)

INTRODUCTION

High-level pattern recognition is extremely difficult, due to the combinatorially explosive growth in the number of possibilities that must be considered. Nevertheless computational processes can always be described in terms of pattern recognition operations. In the case of conventional digital computers these are simple, standard operations acting on precisely defined signal patterns (e.g. NAND gates). Biological cells and organisms, by contrast, are controlled by a vast diversity of extremely powerful molecular pattern recognizers that, roughly speaking, act on the basis of lock–key-type interactions.

These intrinsic recognition processes make a direct as well as supportive contribution to the computational capabilities of cells and organisms. We illustrate this with a conceptual pattern recognition device, previously referred to as the self-assembly model[1,2], and then consider proposed instantiations based on nucleic acids and proteins.

SELF-ASSEMBLY MODEL

The operative principle of the self-assembly device is schematically indicated in *Figure 1*. Symbolic inputs are first converted to molecular shapes that self-assemble like the pieces of a jigsaw puzzle. Different shape features of the self-assembled complex are associated with different groupings of the input patterns. Readout enzymes then determine which shape features are used to trigger output actions. The system thereby converts a symbolic pattern recognition problem to a free energy minimization process. All the nuclear and electronic interactions that contribute to self-assembly are brought to bear on pattern processing at the device level.

Note that reaction kinetics involving substrates are not required for the recognition process to occur. The contribution of conformational pattern recognition is direct in the following sense: it would not be possible to adequately understand or account for the input–output transform performed by the module as a whole by characterizing the functionality of the recognizing molecules in terms of highly compressed descriptions (i.e. a reasonably small set of parameters) which render consideration of the physical interactions among them irrelevant[3]. Of course elaborations of the model can be considered in which reaction–diffusion dynamics also plays a role, either by coupling different molecular recognition processes or by adding another level of pattern processing activity.

Models can be constructed in which quantum features are pertinent, due to the interaction between nuclear and electronic degrees of freedom[4]. Delocalized electrons, such as electrons that tunnel through hydrogen bond pathways, are inevitably in superpositions of different energy states. The resulting interference effects open up new reaction pathways (e.g. by agitating hydrogen bonds). The importance of such effects is at present unknown; but in principle, they allow the parallelism inherent in the quantum mechanical wave function to percolate up to the macro level of function[5,6].

The rate of processing in the device as described above would be diffusion limited. However, release and diffusion of the molecular shapes are not required. A single complex or even a single macromolecule could use its conformational dynamics to fuse patterns of milieu features (*Figure 2*).

TOWARDS IMPLEMENTATION

Figure 3 illustrates a proposed implementation using DNA[7]. The design is based on the fact that methylation influences the equilibrium of B and Z DNA (left- and right-handed forms)[8]. In the variant illustrated a DNA backbone is set up to which complementary oligonucleotides can be attached. Input signals (represented by 0's and 1's)

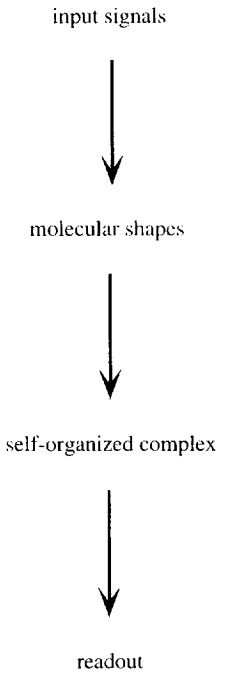

Figure 1 Schematic of self-assembly computing

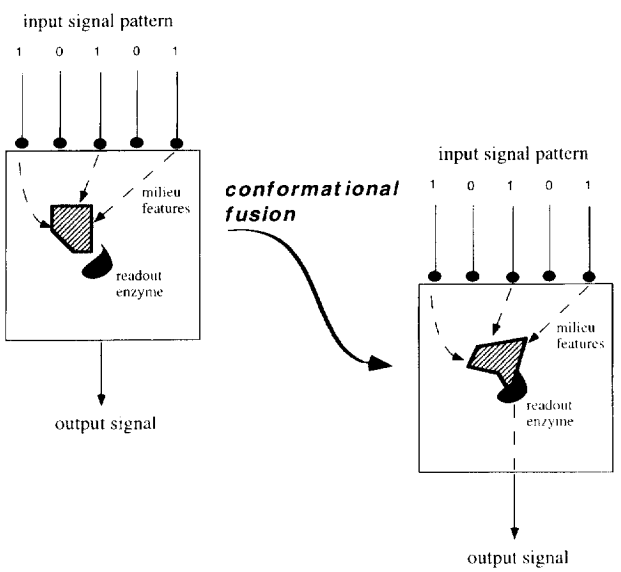

Figure 2 Conformational processor. The input signal pattern is re-coded as a pattern of milieu influences. These are fused by the conformation dynamics of a macromolecular component to a variety of recognizable shape features, each associated with a different family of input patterns. The shape features are connected to output actions by readout enzymes. The macromolecular component could be a single macromolecule (e.g. a single protein) or a polymacromolecular complex

arriving along different input lines are coded into two sets of corresponding oligonucleotides, each with a specific base sequence that can attach to a complementary sequence on the backbone. The oligonucleotides in one of the corresponding sets are unmethylated and those in the other are methylated. The unmethylated member of a corresponding pair is released when the input signal is 0, while the methylated member is released when the signal is 1. If an

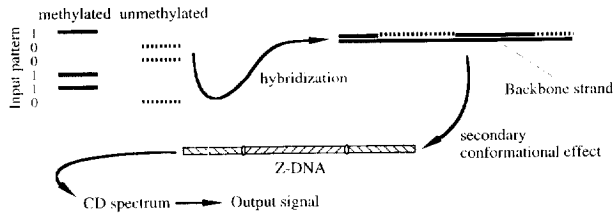

Figure 3 DNA hybridization scheme illustrating the self-assembly concept of computing. Input signal lines carrying a pulse (1) release methylated DNA strands, while dormant lines (0) are encoded as unmethylated strands

input signal line is in a 1 state a corresponding oligonucleotide is methylated; otherwise it is unmethylated. The input signal pattern is thereby re-represented as a pattern of unmethylated and methylated oligonucleotides. The resulting conformational effect (i.e. the B–Z equilibrium) is read out with a circular dichroism (CD) spectrometer[9].

The above design might be called semiprogrammable. The hybrid products can be predicted on the basis of simple rules. But the main part of the computation is performed by the conformational dynamics that occurs subsequent to hybridization. Evolutionary methods of adapting the system are appropriate, since it would not be easy to know in advance which oligonucleotide sequences would lead to a useful grouping of the input patterns. This is actually easier, at this point in time, with DNA than with proteins, since much less effort is required to generate different DNA sequences from oligonucleotide building blocks using hybridization and ligase reactions than to synthesize variant amino acid sequences.

The potential of the DNA conformational processor resides in the great fan-in of input that is possible (between 10^3 and 10^4 input lines for reasonable choices of reaction conditions[10]). The interaction strengths of different signal lines can be altered by altering the arrangement of sequences on the backbone. Nevertheless, the linear character of DNA implies limits in this respect. The use of CD readout, though convenient, would have the disadvantage that detailed conformational effects would be lost.

Similar designs can be constructed on the basis of catalytically active RNA. This offers the possibility of combining the advantages of complex formation based on complementary strands with enriched conformational dynamics based on folded shape. The set of realizable input–output mappings is increased, but at the expense of reduced signal fan-in.

Proteins offer the conformationally richest dynamics, connected with the fact that shape interactions are more effective in an intrinsically 3-D (e.g. globular) form. Here the approach is to choose a naturally occurring macromolecule or complex and then to determine how its activity depends on different combinations of milieu features. Input signals would then be coded into these features. At this stage one would already have a primitive functional module. The module could initially be adapted for desired functions by applying variation-selection operations to the milieu features. More specific molding would be possible if these operations

Table 1 Probable comparative potentialities of DNA, RNA and protein. The ordering is intended to be suggestive over a range of possible computational goals and could be altered by technological advances

	DNA	RNA	Protein
Signal fusion	Lowest	Intermediate	Highest
Evolutionary flexibility	Lowest	Intermediate	Highest
Conformational flexibility	Lowest	Intermediate	Highest
Programmability	Highest	Intermediate	Lowest
Reaction coupling	Lowest	High	High
Output coupling	Lowest	Intermediate	Highest
Current lab convenience:			
Directed evolution	Highest	Intermediate	Intermediate
Handling (stability, protocols)	High	Lowest	Wide range

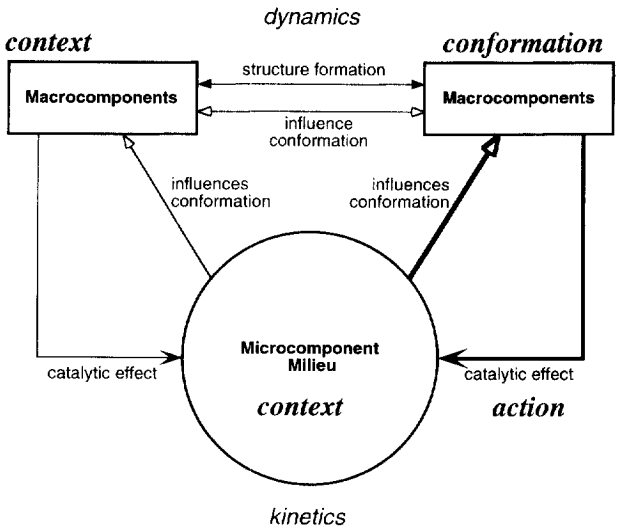

Figure 4 Circular (feedback) relation between context, conformation, and action in complex biochemical networks. The context experienced by a macrocomponent is provided by the microcomponent milieu and by other macrocomponents in its immediate neighborhood. This context determines the conformational state of the macrocomponent, and therefore its catalytic effect on the milieu components and also its shape interactions with macrocomponent neighbors (potentially altering larger-scale structures). The macro-component's action is thus determined by its context and at the same time affects the context of other macrocomponents and of itself

were applied to the proteins per se. As indicated above this would at present be technically more difficult than with DNA, but it should become increasingly feasible as the technology of directed protein evolution advances.

The protein conformational processor (schematically illustrated in *Figure 2*) affords increased possibilities for coupling the conformational dynamics to reaction processes that can be used for readout or for cascading multiple conformational recognition stages. Readout could employ spectroscopic detection of reaction products. Biosensor or electrochemical methods of readout are also available.

Protein-based processors are less programmable than the DNA or RNA variants, since no simple rules are available for determining which complexes will actually form[11]. This is the reason for starting with an existing complex and manipulating the coding of the input signals, at least in the first step. The semiprogrammability of the DNA and RNA processors actually entails a limitation on their computational capabilities, since it means that a large number of interactions that could potentially contribute to the recognition process are frozen out.

Conformational processing devices could be used as molecular co-processors that provide pattern recognition functionality to conventional computing systems. Networks of conformation-driven devices with different capabilities would be well suited to real-time recognition, effector control and data fusion tasks (see *Table 1*). Hybrid designs that work with various combinations of DNA, RNA, and protein would make it possible to integrate the specific advantages of these different materials (e.g. the fan-in of DNA, the context sensitivity of proteins, and the coupling capabilities of RNA). Clearly biological nature has chosen to work in this manner.

COMPLEX CONFORMATIONAL NETWORKS

Biological cells are vastly more complex than the simple designs sketched above. Kinetic, structural, and conformational factors interact in a highly circular manner (*Figure 4*). Conformational interactions largely determine both the kinetic and structural organization, the former through enzyme catalysis and the latter through self-assembly. Molecular conformations, in turn, are influenced by the resulting milieu features. To the extent that the controlling molecular recognizers can be characterized by a small number of definite parameters, their contribution to cell level pattern processing capabilities can be reasonably viewed as supportive only. To the extent that it is in principle necessary to consider the physical dynamics of the recognizers the contribution is direct. The conceptual self-assembly model and proposed implementations considered here suggest that the direct contribution could be quite significant.

Investigating the contribution of such complicated interactions to the information and control capabilities of biological cells is difficult. Computer simulation tools can facilitate the design of meaningful experiments. The CKSD simulation tool described in the accompanying paper (along with a sample run) has been developed with this in mind[12]. The acronym stands for conformation, kinetics, structure, and dynamics. The CKSD system was designed to study the circular relationships depicted in *Figure 4*, both for the purposes of analyzing intracellular dynamics and for designing artificial devices.

The pertinent point here is that the circular (feedback loop) relationship among conformation, kinetics, and structure formation carries with it a natural adaptive character. If the input–output transform performed by the cell is inappropriate the resulting error signal should lead to structural reorganizations and therefore to an altered input–output transform, and so on, until a suitable behavior is discovered. This is a qualitatively broader control capability than that exhibited by today's technological systems, where the physical structure of the components is essentially fixed. For example, a thermostat may control the heat output of a boiler; but it does not

rebuild the boiler to desired specifications. We anticipate that the extended adaptive capability found in biological systems, where feedback can act on structure, could also be exploited in artificial conformational processors.

ACKNOWLEDGEMENT

This research was supported by NSF Grant No. ECS-9704190.

REFERENCES

1 Conrad, M. in '*Advances in Computers*', (Ed. M.C. Yovits), Vol. 31, Academic Press, Inc., Boston, 1990, p. 235–324
2 Conrad, M. *Computer (IEEE)* 1992, **25**(11), 11–20
3 Conrad, M. *BioSystems*, 1995, **35**, 157–160
4 Conrad, M. *Int. J. Quantum Chem.: Quantum Biology Symp.* 1992, **19**, 125–143
5 Conrad, M. *Nanobiol.* 1993, **2**, 5–30
6 Conrad, M. *FED J. (Japan Research & Development Association for Future Electron Devices)* 1995, **6** (suppl. 2), 46–60
7 Conrad, M. and Zauner, K.-P. *BioSystems* 1998, **45**, 59–66
8 Zacharias, W. in '*DNA Methylation: Molecular Biology and Biological Significance*', (Eds J.P. Jost and H.P. Saluz) Birkhäuser Verlag, Basel, 1993, p. 27–38
9 Kennard, O. and Salisbury, S.A. in '*Molecular Biology and Biotechnology*' (Ed. R.A. Meyer), VCH Publishers (Verlag Chemie), Weinheim, 1995, p. 242–247
10 Harwood, A.J. '*Methods in Molecular Biology*', Vol. 58, Humana Press, Totowa, 1996
11 Conrad, M. *Commun. ACM*, 1985, **28**, 464–479
12 Conrad, M. and Zauner, K.-P. In *Computer Science and Biology—Proceedings of the German Conference on Bioinformatics (GCB'96)* (Eds R. Hofestädt, T. Lengauer, M. Löffler, and D. Schomburg), Vol. 1278 of Lecture Notes in Computer Science, Leipzig, 1996. Springer, Berlin, p. 1–10